 doi:10.20944/preprints202011.0649.v1

Article

BFRVSR: A Bidirectional Frame Recurrent Method for Video Super-Resolution

Xiongxiong Xue[1,2], Zhenqi Han[2], Weiqin Tong[1] , Mingqi Li[2], ,and Lizhuang Liu[2,*]

[1] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; xuexiongxiong2018@sari.ac.cn (X.X.); wqtong@shu.edu.cn (W.T.)
[2] Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China; hanzq@sari.ac.cn (Z.H.), limq@sari.ac.cn (M.L.);
* Correspondence: liulz@sari.ac.cn (L.L.)

Abstract: Video super-resolution, which utilizes the relevant information of several low-resolution frames to generate high-resolution images, is a challenging task. One possible solution called sliding window method tries to divide the generation of high-resolution video sequences into independent sub-tasks, and only adjacent low-resolution images are used to estimate the high-resolution version of the central low-resolution image. Another popular method named recurrent algorithm proposes to utilize not only the low-resolution images but also the generated high-resolution images of previous frames to generate the high-resolution image. However, both methods have some unavoidable disadvantages. The former one usually leads to bad temporal consistency and requires higher computational cost while the latter method always can not make full use of information contained by optical flow or any other calculated features. Thus more investigations need to be done to explore the balance between these two methods. In this work, a bidirectional frame recurrent video super-resolution method is proposed. To be specific, a reverse training is proposed that the generated high-resolution frame is also utilized to help estimate the high-resolution version of the former frame. With the contribution of reverse training and the forward training, the idea of bidirectional recurrent method not only guarantees the temporal consistency but also make full use of the adjacent information due to the bidirectional training operation while the computational cost is acceptable. Experimental results demonstrate that the bidirectional super-resolution framework gives remarkable performance that it solves the time-related problems when the generated high-resolution image is impressive compared with recurrent-based video super-resolution method.

Keywords: video super-resolution; bidirectional; recurrent method; sliding window method

1. Introduction

Video super resolution, which solves the problem of reconstructing high-resolution images from low-resolution images, is a classic problem in image processing. It is widely used in security, entertainment, video transmission and other fields. Compared with single image super-resolution, video super-resolution can use more information to output better high-resolution images, such as feature information of adjacent frames. However, the reconstruction of video super resolution images is generally difficult because of various issues, such as occlusion, adjacent frame information utilization and computational cost.

With the rise of deep learning, video super-resolution has received significant attention from the research community over the past few years. Sliding window method and recurrent method are two latest state-of-the-art methods based on deep learning. Specifically, Sliding Window Video Super-Resolution (SWVSR) solves this problem by combining a batch of low-resolution images to reconstruct a single high-resolution frame, and divides the video super-resolution task into multiple independent super-resolution subtasks [1]. Each input frame will be processed several times, which will cause waste of calculations. In addition, the generation process is an independent sub-task, which may reduce time consistency, resulting in flickering and artifacts. Unlike SWVSR, Recurrent Video

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2020 doi:10.20944/preprints202011.0649.v1

Super-Resolution (RVSR) generates the current high-resolution image from the previous high-resolution image, the previous low-resolution image and the current low-resolution image [2, 3]. Each input frame will be processed one time. RVSR has the ability to process video sequences of any length, and enables the details of the video to be implicitly transmitted in longer video sequences. Insufficient use of information caused by RVSR leads to correlation between image quality and time (Show in Fig.1).

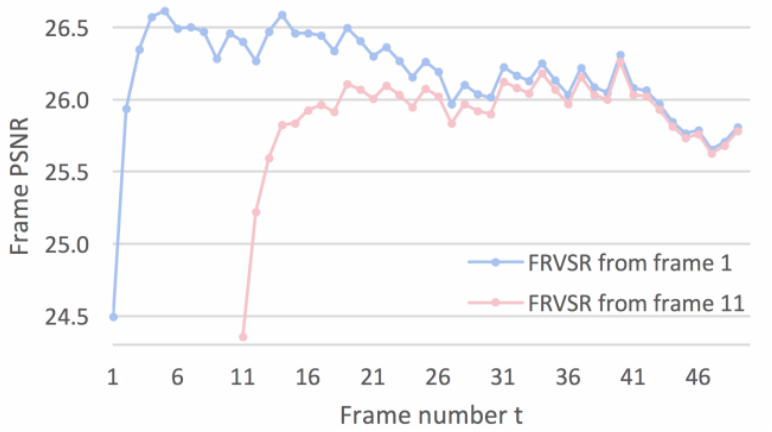

Figure 1. Picture from FRVSR [3]. FRVSR is state-of-the-art method of RVSR. RVSR can handle video sequences of any length, but there are some problems. As shown in the figure, the RVSR method has a problem that is correlation between image quality and time.

In our work, we propose an end-to-end trainable Bidirectional Frame Recurrent Video Super-Resolution (BFRVSR) framework to address the above issues. We adopt the forward training and the reverse training to solve the problem of insufficient utilization of information and preserve temporal consistency shown in Fig.2. BFRVSR has several benefits, which gains the balance between RVSR and SWVSR. Each input frame needs to be processed no more than twice while each output frame makes full use of the information contained by optical flow or any other calculated features. In addition, passing the previous high-resolution estimate directly to the other step helps the model to recreate fine details and produce temporally consistent videos.

Our contribution is mainly reflected in the following: a.） Propose a Bidirectional Frame Recurrent Video Super-Resolution framework. b.） An end-to-end video super-resolution model based on Bidirectional Frame Recurrent Video Super-Resolution framework is proposed, and no pre-training step is required. c.） Address the correlation between image quality and time and preserve temporal consistency.

2. Releted Work

Image Super-Resolution (ISR) is a classic ill-posed problem. The methods are divided into interpolation methods, such as nearest, bilinear, bicubic, and dictionary learning [4, 5], example-based methods [6, 7, 8, 9, 10], and self-similarity approaches [11-14]. We refer the reader to three review documents [15-17] for extensive overviews of prior art up to recent years.

The recent progress in deep learning, especially in convolutional neural networks, has shaken up the field of ISR. Single Image Super-Resolution (SISR) and Video Super-Resolution are two categories based on ISR.

SISR uses a single low-resolution image to estimate a high-resolution image. Dong et al. [18] introduced deep learning into the field of super-resolution. They imitated the classic super-resolution solution method and proposed three steps of feature extraction, feature fusion, and feature reconstruction to complete SISR. Then, K. Zhang et al. [19] reached state-of-the-art results with deep CNN networks. A large number of excellent results have emerged [20-24]. Parallel efforts have studied the loss function [25-27].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2020 doi:10.20944/preprints202011.0649.v1

Video super-resolution combine information from multiple LR frames to reconstruct a single high-resolution frame. Sliding window method and recurrent method are two latest state-of-the-art methods.

Sliding window method divide the video super-resolution task into multiple independent subtasks, and each subtask generates a single high-resolution output frame from multiple low-resolution input frames [1, 28-30]. The input is adjacent 2N+1 frames of low-resolution images like $\{I_{t-N}^{LR}, I_{t-N+1}^{LR} \dots, I_t^{LR}, \dots I_{t+N-1}^{LR}, I_{t+N}^{LR}\}$. Then, an alignment module is used to align $\{I_{t-N}^{LR}, I_{t-N+1}^{LR} \dots I_{t+N-1}^{LR}, I_{t+N}^{LR}\}$ with the I_t^{LR}. Finally, I_t^{HR} is estimated through the aligned *2N+1* low-resolution frames. Drulea and Nedevschi *et al.* [23] used optical flow method to align I_{t-1}^{LR} and I_{t+1}^{LR} with I_t^{LR} and use them to estimate I_t^{HR}.

Recurrent method generates a high-resolution image from the previous high-resolution image, the previous low-resolution image and the low-resolution image. Huang *et al.* [31] used a bidirectional recurrent architecture, but did not use any explicit motion compensation in their model. Recurrent structures are also used for other tasks, such as blurring [32] and stylization[33, 34] of videos. Kim et al. [32] and Chen et al. [33] passed the feature representation to the next step, and Gupta et al. [34] passed the previous output frame to the next step, generating time-consistent stylizations in parallel work video. Sajjadi et al. [3] proposed a recursive algorithm for video super-resolution. The FRVSR [3] network estimates the optical flow $F_{t\to t-1}^{LR}$ of I_{t-1}^{LR} and I_t^{LR}, and uses I_{t-1}^{HR} and $F_{t\to t-1}^{LR}$ to generate $\tilde{I}_t^{HR}$. Finally, sends $\tilde{I}_t^{HR}$ and I_t^{LR} to the network for reconstruction to obtain I_t^{HR}. However, insufficient use of information caused by FRVSR leads to correlation between image quality and time.

3. Methods

The framework of BFRVSR is shown in the Fig.2. All network modules can be replaced. For example, the optical flow module can use existing methods that have been pre-trained instead of training and building the network from scratch. You can also consider using a deformable convolution module [35] to replace the optical flow module.

After presenting an overview of the BFRVSR framework in Sec. 3.1 and defining the loss functions used for training in Sec. 3.2, we justify our design choices in Sec. 3.3.

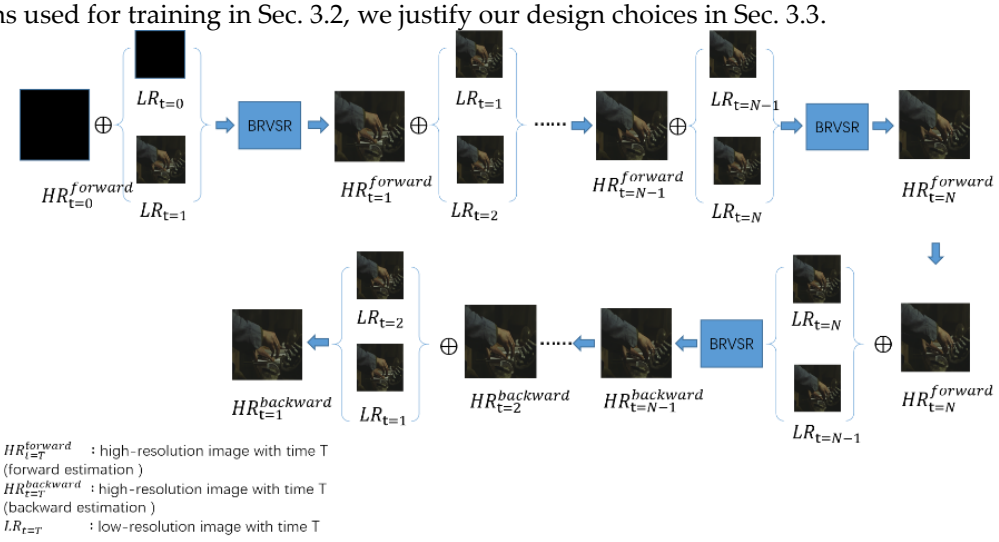

Figure 2. Overview of the proposed BFRVSR framework. BFRVSR not only performs the forward operation of the RVSR, but also re-estimates { $HR_{t=1}^{backward}$, $HR_{t=2}^{backward}$ …… $HR_{t=N-2}^{backward}$, $HR_{t=N-1}^{backward}$} through the generated $HR_{t=N}^{forward}$. BFRVSR ensures that the generation of all High resolution image frames depends on global information, rather than local information.

3.1. BFRVSR

The proposed model is shown in Fig.3. Trainable modules include optical flow estimation network FlowNet and super-resolution network SuperNet. The input of our model is the low-resolution image of the current frame I_t^{LR}, the low-resolution image of the previous frame I_{t-1}^{LR}, and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2020 doi:10.20944/preprints202011.0649.v1

the high-resolution image estimation of the previous frame I_{t-1}^{HR}.The output of our model is the high-resolution image estimation of the previous frame I_t^{HR}.

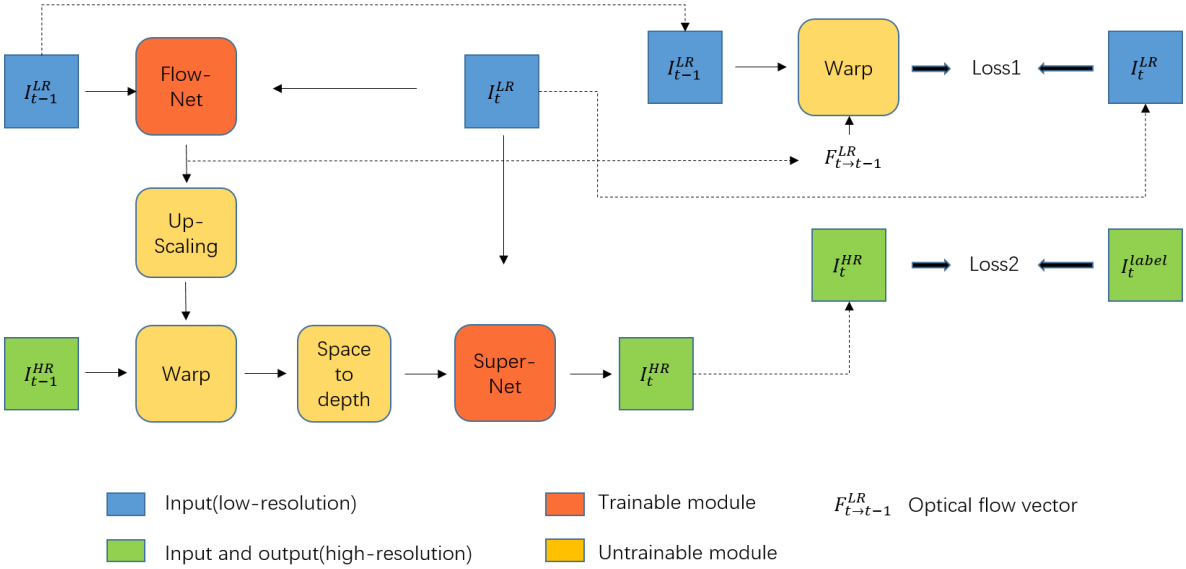

Figure 3. Overview of training network framework(Left). Trainable modules include FlowNet and SuperNet. Up-sampling uses bilinear interpolation. Loss function used during training(right).

Flow estimation： The network structure of FlowNet [3] is shown in the Fig.4. Firstly, the network uses the optical flow estimation module to estimate the low-resolution image of the previous frame I_{t-1}^{LR} and the low-resolution image of the current frame I_t^{LR} to obtain a low-resolution motion vector diagram $F_{t\to t-1}^{LR}$.

$$F_{t\to t-1}^{LR} = FlowNet(I_{t-1}^{LR} \oplus I_t^{LR}) \in [-1,1]^{H\times W\times 2} \quad (1)$$

$F_{t\to t-1}^{LR}$ shows the position information from the current image to the previous frame.

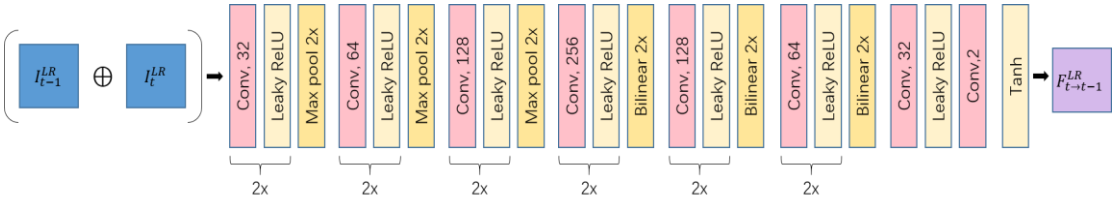

Figure 4. Overview of FlowNet. 2x represents the linear superposition of two identical modules. The input is { I_{t-1}^{LR} , I_t^{LR}}, and the output is $F_{t\to t-1}^{LR}$ through FlowNet module. $F_{t\to t-1}^{LR}$ represents the x displacement vector and y displacement vector between I_t^{LR} and I_{t-1}^{LR}.

Upscaling flow: In this step, we process the low-resolution optical flow map that has been obtained, and we use bilinear interpolation with scaling factor s to up-sampling to obtain the high-resolution optical flow map.

$$F_{t\to t-1}^{HR} = Upsample(F_{t\to t-1}^{LR}) \in [-1,1]^{sH\times sW\times 2} \quad (2)$$

Warping HR image: Use the obtained high-resolution optical flow diagram and the high-resolution image of the previous frame to estimate the high-resolution image of the current frame.

$$\tilde{I}_t^{HR} = Warp(I_{t-1}^{HR}, F_{t\to t-1}^{HR}) \quad (3)$$

We implemented warping as a differentiable function using bilinear interpolation similar to Jaderberg et al. [36].

Mapping to LR space: We map high-dimensional spatial information to low-dimensional depth information using the space-to-depth transformation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2020 doi:10.20944/preprints202011.0649.v1

$$H_t^{depth} = DM(\tilde{I}_t^{HR}) \tag{4}$$

Our method of mapping to low-dimensional space is similar to the method in FRVSR [3]. The Mapping to LR space operation process is shown in the Fig.5.

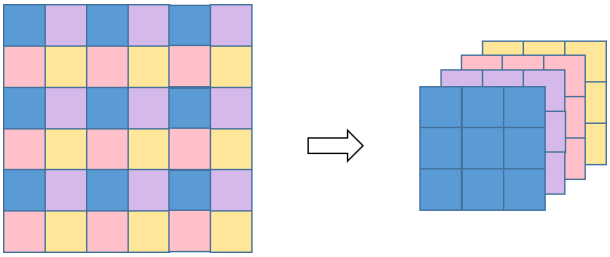

Figure 5. Space-to-Depth module. Compress the spatial information of high-resolution images into low-resolution image depth information.

Super-Resolution: In this step, the low-dimensional depth map of the high-resolution image of the current frame H_t^{depth} and the low-resolution image of the current frame I_t^{LR} are sent to the SuperNet to obtain the final high-resolution frame. The network structure of SuperNet is shown in the Fig.6.

$$I_t^{HR} = SuperNet\left(H_t^{depth} \oplus I_t^{LR}\right) \tag{5}$$

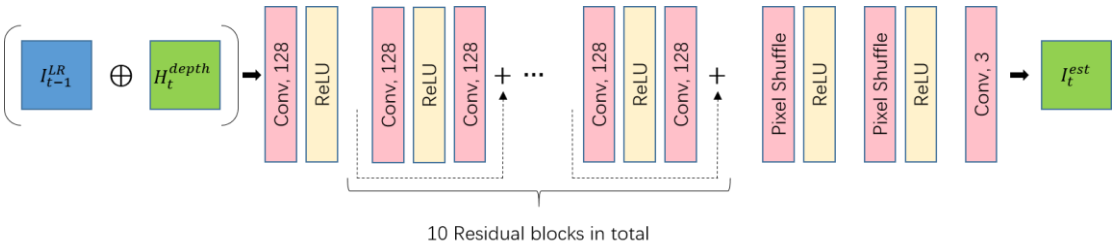

Figure 6. Overview of SuperNet. SuperNet uses the RESNET framework and Pixel Shuffle upsampling operation. SuperNet is an open framework and can be replaced by other networks.

Summary: The overall process of the network is as follows:

$$I_t^{HR} = SuperNet(DM(Warp(I_{t-1}^{HR}, Upsample(FlowNet(I_{t-1}^{LR} \oplus I_t^{LR}))) \oplus I_t^{LR}) \tag{6}$$

3.2. Loss functions

In our network architecture, the optical flow estimation module and the super-resolution module are trainable, so in the training process, two loss functions are used to optimize the results.

The first loss function is the error between the high-resolution image generated by the super-resolution module and the real image label I_t^{lable}.

$$L_1 = ||I_t^{HR} - I_t^{label}||_2^2 \tag{7}$$

Because the data set does not have the ground truth of optical flow, we use a method similar to the FRVSR [3] to calculate the spatial mean square error on the curved LR input frame to optimize the optical flow estimation module as the second loss function.

$$L_2 = ||Warp(I_{t-1}^{LR}, F_{t\to t-1}^{LR}) - I_t^{LR}||_2^2 \tag{8}$$

The Loss function of training final backpropagation is $L_{total} = L_1 + L_2$.

3.3. Justifications

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2020 doi:10.20944/preprints202011.0649.v1

The motivation for proposing the BFRVSR framework is as follows：

a.） The super-resolution network using the sliding window method has high computational cost. Each frame of image needs to be calculated 2N+1 times (window size 2N + 1). We use the bidirectional frame recurrent to process each frame at most twice.

b.）Direct access to the output of the previous frame can help the network generate a temporally consistent estimate for the next frame or previous frame. In the recurrent network, insufficient use of information leads to correlation between image quality and time. Through bidirectional network operation, each frame of image has all frames information in the window.

4. Experiment

4.1. Training Datasets and Details

Training datasets. Vimeo-90k [37] is our training and testing data set. We abbreviate the Vimeo-90k test data set as Vimeo-90k-T and the Vimeo-90k train data set as Vimeo-90k-TD. Vimeo-90k data set contains 91701 7-frames continuous image sequences, and is divided into Vimeo-90k-TD and Vimeo-90k-T. In the training data set, we randomly crop the original 448×256 image into a 256×256 real label image. In order to generate LR images, we perform Gaussian blur and down-sampling processing on the real label image, and use a Gaussian blur with standard deviation $\sigma = 2.0$.

Training details. Our network is end-to-end trainable, and there are no modules that need to be pre-trained. The Xavier method is used for initialization. We train 600 epochs and the batch size is 4, the optimizer uses Adam optimizer, and the initial learning rate is 10^{-4}, which is reduced by 0.1 times every 100 epochs. In a batch, each sample is 7 consecutive images. We conduct video super-resolution experiments at 4x factor.

In order to obtain the first high-resolution image I_1^{HR}, two methods can be used. In the first method, we set I_0^{HR} to a completely black image. This can force the network to learn detailed information from low-resolution images. In the second method, we upsample I_1^{LR} to I_1^{HR} through the bicubic interpolation method, and estimate I_2^{HR} from $\{I_2^{LR}, I_1^{LR}, I_1^{HR}\}$.In order to compare with the RVSR method, we used the first method for experimentation.

4.2. baselines

For a fair evaluation of the proposed framework on equal ground, we compare our model with three baselines that use the same optical flow and super-resolution networks.

SISR：Only a single low-resolution image is used to estimate a high-resolution image without relying on timing information. The input is I_t^{LR} and the output is I_t^{HR}.

VSR：Through$\{I_{t-1}^{HR}, I_{t-1}^{LR}, I_t^{LR}\}$, without the optical flow network estimation, relying on the learning space deformation ability of the convolution operation itself to obtain I_t^{HR}.

RVSR：Through $\{I_{t-1}^{HR}, I_{t-1}^{LR}, I_t^{LR}\}$, with the optical flow network estimation, and then sent to SuperNet to obtain I_t^{HR}. The operation process is the same as the forward propagation in the BFRVSR network.

We ensure that the network model is consistent during the evaluation. The key parameters of the training parameters are the same. The initialization uses Xavier initialization, and the accelerator uses Adam optimizer. The initial learning rate is 10e-4, which is reduced to 0.1 times every 100 rounds. All networks are trained with the same training set, and the coefficient of Gaussian blur is 2.0.

4.3. Analysis

We train baselines and BFRVSR to convergence under the same parameter conditions. We compare and test the pre-trained model on the Vimeo-90K-T. Table 1 shows the comparison image PSNR results of baselines and BFRVSR. Compared with baselines, our proposed framework has the best effect in continuous 7-frames video sequences, and it is 0.39dB higher than the RVSR method. PSNR of BICUBIC and SISR is only related to current low-resolution images, and no correlation between high-resolution images. PSNR of VSR and RVSR has correlation between image quality and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 November 2020 doi:10.20944/preprints202011.0649.v1

time. Because of motion compensation by optical flow network, RVSR performance is better than VSR.

Table 1. The PNSR index of the image generated by the five methods of BFRVSR, RVSR, VSR, SISR, and BICUBIC are compared. As can be seen in the figure, BFRVSR is an upgrade of RVSR, which not only has the best effect, but also overcomes the shortcomings of RVSR's unidirectional gain.

	Frame1	Frame2	Frame3	Frame4	Frame5	Frame6	Frame7	Average
BICUBIC	29.3057	27.3187	29.3173	29.3120	27.3087	27.3051	27.2900	27.3082
SISR	28.5332	28.5633	28.5240	28.5468	28.5523	28.5447	28.5593	28.5462
VSR	28.7632	29.4320	29.8012	29.8122	29.8310	29.9001	29.9212	29.6373
RVSR	29.0803	29.8807	30.1547	30.2898	30.3553	30.3980	30.3991	30.0797
BFRVSR(ours)	**30.4772**	**30.4836**	**30.4833**	**30.4739**	**30.4670**	**30.4547**	**30.4145**	**30.4649**

BFRVSR will perform a forward estimation and a reverse estimation. BRVSR is equivalent to an RVSR network in forward estimation. It transmits global detail information by using I_{t-1}^{HR} and performs timing alignment operations. However, there are some problems, that is, the details of I_j^{LR} cannot be obtained for I_i^{LR} to optimize the image (i > j). Reverse estimation solves this problem. Reverse estimation makes each frame implicitly use all the information to estimate the high-resolution image of the frame. Use the $\{I_t^{HR},\ I_{t-1}^{LR},\ I_t^{LR}\}$ to generate I_{t-1}^{HR}.

RVSR can be trained on video clips of any length. However, if the video clip is too long, RVSR will have a problem that is correlation between image quality and time. In fact, RVSR also has the problem on shorter video clips. BFRVSR solves this problem as shown in Fig.7.

The video super-resolution based on the sliding window method processes each frame 2N+1 times, the video super-resolution based on the recurrent method processes each frame once, and the BFRVSR processes each frame at most 2 times.

On the RTX-2080Ti, the time for a single image Full HD frame for 4x super-resolution is 291 ms.

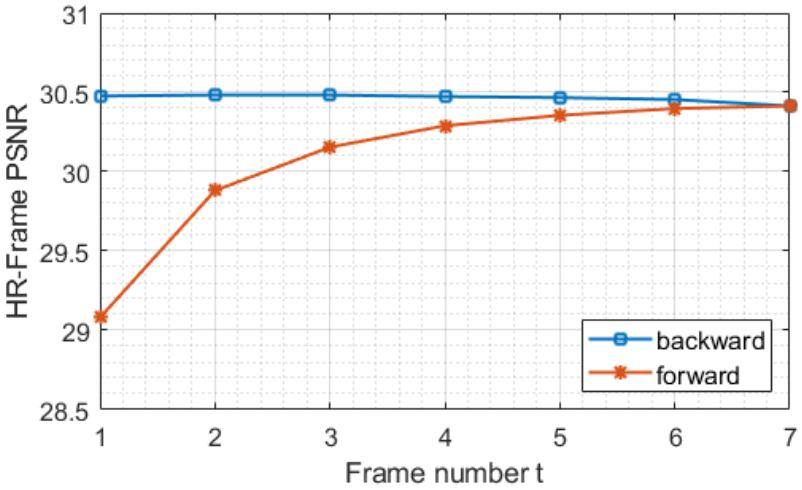

Figure 7. We show the quality of each frame in the forward propagation of BFRVSR and the quality of each frame in the reverse propagation. We show the quality of each frame in the forward propagation of BFRVSR and the quality of each frame in the reverse propagation. We found that global information is implicitly used in backpropagation to generate high-resolution images.

5. Conclusions

We propose an end-to-end trainable bidirectional frame recurrent video super-resolution method. Due to the operation of bidirectional training, with more information utilized to feed the model to deal with the correlation between image quality and time, BFRVSR successfully solves the problem happened in Fig.1, to be specific, it decouples the correlation between image quality and time. In addition, the proposed method achieves better image quality while the computational cost

is lower than sliding window method. Nevertheless, there is still room for improvement in the field of video super-resolution. On the one hand, if the problem of occlusion and blur is considered, much more computational cost would be required. We may deal with the problem by adding cross connections. On the other hand, a deformable convolution module, which has been frequently investigated recently, shows enormous potential in the field of image classification, semantic segmentation, etc. Thus it may achieve better results if we replace the optical flow module with deformable convolution module. Furthermore, it is believed that video super-resolution and frame insertion have considerable similarities thus we may try to utilize BFRVSR to perform these two tasks simultaneously.

Author Contributions: Project administration, X.X.; Validation, X.X.; investigation, X.X., Z.H.; resources, W.T. ,M.L., L.L.; visualization, X.X, Z.H.

Funding: Supported by the National Natural Science Foundation of China (Grant No. 61972007).

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Wang X , Chan K C K , Yu K , et al. EDVR: Video Restoration With Enhanced Deformable Convolutional Networks[C]. In Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Angeles, USA,16-19 June,2019.
2. Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In CVPR, 2017.
3. Sajjadi M S M , Vemulapalli R , Brown M . Frame-Recurrent Video Super-Resolution[C]// In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 18-22 June, 2018.
4. J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing*, **2012**.
5. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, ´ D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition , Las Vegas , USA, 26 June – 1 July,**2016**.
6. C. E. Duchon. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology*, **1979**.
7. G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Transactions on Graphics*, **2011**.
8. W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, **2002**.
9. R. Timofte, R. Rothe, and L. Van Gool. Seven ways to improve example-based single image super resolution. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition , Las Vegas , USA, 26 June – 1 July,2016.
10. J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, Oregon, USA, 23-28 June, 2013.
11. C. Liu and D. Sun. A bayesian approach to adaptive video super resolution. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA, 21-25 June, 2011.
12. J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In CVPR, 2015.
13. O. Makansi, E. Ilg, and T. Brox. End-to-end learning of video super-resolution with motion compensation. In Proceedings of GLOBAL CONFERENCE on PSYCHOLOGY RESEARCHES, Lara–Antalya, Turkey, 16-18 March, 2017.
14. A. Ranjan and M. J. Black. Optical flow estimation using a spatial pyramid network. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA,21-26 July,2017.
15. Anwar S , Khan S , Barnes N . A Deep Journey into Super-resolution: A survey[J]. *arXiv* **2019** ,arXiv:1904.07523.
16. Wang Z , Chen J , Hoi S C H . Deep Learning for Image Super-resolution: A Survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2020**, PP(99):1-1.

17. K. Nasrollahi and T. B. Moeslund. Super-resolution: A comprehensive survey. *Machine Vision and Applications,* **2014**.
18. C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In ECCV, 2014.
19. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising" In Proceedings of IEEE Transactions on Image Processing, 2017.
20. W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang. Deep laplacian pyramid networks for fast and accurate superresolution. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA,21-26 July,2017.
21. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Te- ´ jani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA,21-26 July,2017.
22. E. Perez-Pellitero, J. Salvador, J. Ruiz-Hidalgo, and B. Rosenhahn. PSyCo: Manifold span reduction for super resolution. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition , Las Vegas , USA, 26 June – 1 July,2016..
23. M. Drulea and S. Nedevschi. Total variation regularization of local-global optical flow. In Proceedings of International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA, 5-7 October,2011.
24. X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia. Detail-revealing deep video super-resolution. In Proceedings of IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October, 2017.
25. C.-Y. Yang, J.-B. Huang, and M.-H. Yang. Exploiting selfsimilarities for single frame super-resolution. In Proceedings of Asian Conference on Computer Vision, 2010.
26. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for realtime style transfer and super-resolution," In Proceedings of European Conference on Computer Vision ,Amsterdam city, Netherlands, 8-16 October, 2016.
27. C. Ledig, L. Theis, F. Husz´ar, J. Caballero, A. Cunningham ,A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang et al., "Photorealistic single image super-resolution using a generative adversarial network," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA,21-26 July,2017.
28. P. Milanfar. Super-resolution Imaging. CRC press, **2010**.
29. Dai J , Qi H , Xiong Y , et al. Deformable Convolutional Networks[J]. **2017**.
30. Xiang X , Tian Y , Zhang Y , et al. Zooming Slow-Mo: Fast and Accurate One-Stage Space-Time Video Super-Resolution[J]. **2020**.
31. Y. Huang, W. Wang, and L. Wang. Bidirectional recurrent convolutional networks for multi-frame super-resolution. In Proceedings of Advances in Neural Information Processing Systems, Montreal, Quebec, Canada, 11 – 12 Dec, 2015.
32. T. H. Kim, K. M. Lee, B. Scholkopf, and M. Hirsch. Online ¨ video deblurring via dynamic temporal blending network. In Proceedings of IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October, 2017.
33. D. Chen, J. Liao, L. Yuan, N. Yu, and G. Hua. Coherent online video style transfer. In Proceedings of IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October, 2017.
34. A. Gupta, J. Johnson, A. Alahi, and L. Fei-Fei. Characterizing and improving stability in neural style transfer. In Proceedings of IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October, 2017.
35. Dai J , Qi H , Xiong Y , et al. Deformable Convolutional Networks[J]. *arXiv* **2017 ,**arXiv:1703.06211.
36. M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu. Spatial transformer networks. In Proceedings of Advances in Neural Information Processing Systems, Montreal, Quebec, Canada, 11 – 12 Dec, 2015.
37. Xue T , Chen B , Wu J , et al. Video Enhancement with Task-Oriented Flow[J]. *International Journal of Computer Vision*, **2017**(1).